SEEDS Technology Infusion Study

Second SEEDS Public Workshop

Capability Vision and Technology Infusion Process
Breakout Session Summary

Breakout Agenda

- **Background (15 min)**
  - Overview of SEEDS technology study, ESE vision, etc.
  - Reference material for working session
- **Capability discussion (40 min)**
  - Write down individual thoughts on capabilities
  - Group discussion – highlight one per person
- **Breakout session instructions (5 min) - 2 groups: mission / applications focus**
- **Breakout session (2 hr + 15 min break)**
  - Barriers
  - Related prototypes
  - Trends
  - Capabilities
  - Vision development forum
  - Vision representation
  - Technology infusion barriers
  - Technology infusion strategies

Breakout Results Summary

SEEDS Formulation

- **Good participation…thank you!**
  - 35+ active participants
- **Identified many important elements of (inputs to) a SEEDS vision**
  - 49+ capabilities
  - 12+ specific features
  - 18+ current barriers to reaching ESE goals
  - 26+ prototypes of relevant technologies
  - 32+ technology/science/application trends
  - 5 barriers to technology infusion
  - 2 recommendations on vision representation
- **Details of input and recommendations on following slides**
  - Lightly summarized results
  - Input from "mission" and "applications" focus groups is provided on separate slides where applicable

Capabilities – Group Discussion Results

SEEDS FOCUS AREA

- **What new or improved capabilities are needed in ESE data systems over the next 10 years?**
  1. Access
     - Info not Data
     - Expert guidance for users ("Ask Jeeves"), Interactive access,
     - Better web services, Semantic web, domain semantics,
     - Catalog-to-data seamless access, content-based search
  2. Data Services
     - Tools to identify and resolve missing data
     - Reduce search result set
     - Automated intelligent QA
     - Support seamless data fusion
     - Data models to support emerging data collection (nano-tech, non-gridded)
  3. Distribution
     - Store compressed data objects, subset on compressed file to deliver
     - Subscribe and deliver to where you are (mobile access, eg to PDA)
     - Output multi-formats (format transparent to user)
  4. Tools
     - Building tailored processing chains (on-the-fly)
     - High level data manipulation language for custom data products (from multi sources)
     - Event recognition without reprocessing or new metadata
     - Adapt tools to commercial products, COTS environments
  5. Automated Ops, System "always up"
  6. Design
     - Simplicity, avoid "gratuitous complexity"
     - Alternatives to firewall, security transparency

Barriers

- **What technical problems hinder your work?**
- **Considerations**
  1. Hard to combine disparate data for more sophisticated models, interdisciplinary research, and applications
  2. Cannot obtain data within 45 minutes of overpass to support weather model initialization
  3. Cannot easily access NASA data sources from commercial software products
- **Check**
  - Does your list cover all missing technical capabilities that prevent us from achieving the 2025 ESE vision/goals now?

Barriers - Applications Focus Results

- **What technical problems hinder your work?**
  1. Slow data access- limited bandwidth or data chunks too big
  2. Security issues
     - Awkward implementations of policy
     - Irrational/simplistic policies (e.g., block all access)
     - Immature authentication technologies esp. for distributed systems
     - Data access and services not tailored to different users – "one size fits all" design results in restrictive policies
  3. Rapid change of technology and cost of porting/migrating apps
  4. Disparate data models across various communities
  5. COTS product and infrastructure issues
     - Platform/OS incompatibilities
     - COTS product interdependencies
     - Slow to support new mission data formats
     - COTS products not providing advertised functionality
  6. Poor system designs
     - Generalized solutions do not meet special needs
     - Specialized solutions so not meet general needs
  7. Knowledge gap between Earth science and computer science
     - Need new computational Earth science field?

Barriers - Mission Focus Results

SEEDS FOUNDATION

- **What technical problems hinder your work?**
  1. Security - firewalls, next gen networks (will security hinder data access?)
  2. Management of multiple goals - quick response and equity considerations (competing goals without software to support conflict resolution)
  3. Bandwidth - (point-to-point as well as space-ground); capacity lacking to fuse from multiple sources to satisfy higher-level needs; commercial tech drivers fall exceed NASA need
  4. Technology evolution - align, adapt; too many goals => difficult to satisfy all uses
  5. Automated systems are too manual
  6. Near-line archives inadequate for needs; volumes too high
  7. User model unknowable (changing); want interface to access knowledge; need easier to interface to subset of interest; need to rus user scripts on server side

Related Prototypes

- **What leading-edge prototypes do you think demonstrate capabilities that should be made pervasive in the SEEDS era?**
- **Considerations**
  1. MODster & DODster: Distributed, Decentralized MODIS Data & Services
  2. Standards Framework in Support of Dynamic Assembly of NewDISS Components
  3. Universal Interchange Technology for Earth Science Data and Services (UNITE)

Related Prototypes - Application Focus Results

SEEDS Foundation

- **What leading-edge prototypes do you think demonstrate capabilities that might be made pervasive in the SEEDS era?**
  1. SRB + Data Grid
  2. OGC Web services testbed
  3. ESML
  4. Echo
  5. OpenDAP (DODS) … and all other ESIP activities presented in plenary
  6. Conquest
  7. Resource Description Framework
  8. XML Topic Maps
  9. Grid prototypes
  10. Fortezza cards
  11. RapidFire
  12. D2K data mining package
  13. GloViz @ EDC
  14. Global Broadcast System
  15. Data pools

Related Prototypes - Mission Focus Results

SELDS FOLLOW-ON

- **What leading-edge prototypes do you think demonstrate capabilities that should be made pervasive in the SEEDS era?**
  1. Data Grid, service broker
  2. EOSDIS as a prototype - interdisciplinary definitions, lessons learned, user patterns, examples of tough problems solved
  3. Near-archive data mining
  4. Interview science teams, DAACs, how is SEEDS improved?
  5. NVO - National Virtual Observatory
  6. Microsoft Terraserver has analogous system
  7. REALM - lessons learned form Sam Goward/UM
  8. Simulation modeling in Hollywood
  9. TV News ability to pull video footage for breaking news reports
  10. Commercial digital archives
  11. Sony game playing distributed architecture and visualization

Trends

- **What current trends do you think foretell the most important changes to user requirements and system designs in the SEEDS era?**
- **Considerations**
  1. LAN services → Internet services
  2. Data poor → Data rich (volume & variety)
  3. Central processing → End-user processing
  4. Single discipline science → Multi-disciplinary science/apps
  5. High-latency data for research → Low-latency data for apps (e.g., weather)

Trends - Application Focus Results

- **What current trends do you think foretell the most important changes to user requirements and system designs in the SEEDS era?**
  1. Offline → near line → on-line interactive
  2. Simple data access → value-added services
  3. Focused research community → broad application user communities
  4. Leading edge technology focus → usable technology focus
  5. Dumb sensors → Smart sensors
  6. Single downlink → direct broadcast
  7. Centralized → distributed (cluster computing, grid computing)
  8. Higher end-user computing capability
  9. Single purpose systems → multipurpose systems (research + applications) a la NPOESS
  10. Planned data products → on demand data product generation
  11. Slow networks → fast networks
  12. Wired → Wireless apps

Trends – Mission Focus Results

- **What current trends do you think foretell the most important changes to user requirements and system designs in the SEEDS era?**
  1. Cheaper on-line storage
  2. Bandwidth may get more expensive compared to storage
  3. Processing on-board satellites
  4. Outsourcing services/equipment
  5. Look to commercial service providers (use developing standards to plug-in)
  6. More server-side processing
  7. Distributed (mobile) services
  8. Smaller specialized services (chained together)
  9. Computation: SETI computational model (leverage unused CPU cycles); Sony/game playing models
  10. Visualization: ESE data role in entertainment/real time games (Sony)
  11. Coupled data (real-time stream) with large model outputs
  12. Charging for data
  13. Space technologies slow to evolve compared to intelligent ground technologies

Capabilities (Revisited) - Application Focus Results

- **What new or improved capabilities are needed in ESE data systems over the next 10 years?**
  1. Ability to produce higher level products over distributed systems (e.g., distributed FFT) and other new architectures
  2. Refined security mechanisms
  3. Levels of service on servers to handle broader set of users
  4. Flexible/adaptable interfaces – systems provide value to engage end-users earlier
  5. Automated near-real-time event detection
  6. Faster data ingest
  7. Separate ingest from distribution

Capabilities (Revisited) - Mission Focus Results

- **What new or improved capabilities (not already captured in the group discussion) are needed in ESE data systems over the next 10 years?**
  1. Improved security management
  2. Tools for management of multiple (conflicting) goals
  3. Shipping code
  4. Outsourcing of network services
  5. Optimized mission planning
  6. Get on the Grid

Features

- **What specific new features are needed in ESE data systems over the next 10 years to enable the capabilities you identified?**
- **Considerations**
  1. ArcInfo and IDL plug-ins that provide direct access to DAAC holdings
  2. Near-real-time data delivery of low-quality data subsets (images) over a limited area (implies prioritized vs. FIFO processing)
  3. Ability to order a data product based on the content of another data product (e.g., high-res data over fires detected in low-res data)
- **Check**
  - Does your list address all needed capabilities?

Features - Application Focus Results

SEEDS FOCUS

- **What specific new features are needed in ESE data systems over the next 10 years to enable the capabilities you identified?**
  1. Direct broadcast
  2. Parallel taps into raw or processed data stream to deliver data to special-purpose plug-in application services w/ near-real-time capabilities
  3. Ability to do quality/timeliness trades (note: AVHRR available 2 hrs after overpass)
  4. Prioritization by user or user type
  5. Strong user identification
  6. Resource usage monitoring
  7. General data pools (limited depth pools for user algorithms)
  8. Data debit cards
  9. Intelligent data subsetting (much more than FTP)
  10. Partial data decompression
  11. Data fusion
      - Semantic understanding of the data (esp. resolutions)
      - Swath + Uniform grid + vector
      - Co-registration
      - Enabling third-party services/products that can be vetted by the users
      - COTS tools understand metadata and can make certain automatic transformations/warnings
  12. Better insight into different qualities of similar datasets

Features - Mission Focus Results

- **What specific new features are needed in ESE data systems over the next 10 years to enable the capabilities you identified?**
  1. Knowledge representation, extraction, and access tools
  2. User interfaces much simpler (partner with commercial developers)
  3. Modules that work together
  4. Defining rules for transforming query to next level
  5. Population knowledge repository - discipline-specific level
  6. Classifier of data into knowledge
  7. Ontology/namespace

Vision Representation

SEEDS Formulation

- **How should the vision for ESE data systems be represented?**
- **Considerations**
  1. List of features/benefits
  2. Future user scenarios
  3. Thematic diagram/picture
  4. Goal system architecture
  5. Tell a story (video w/ animation, interviews, etc.)

Vision Representation – Mission + Application Focus

SEEDS FOCUS

- **How should the vision for ESE data systems be represented?**
  1. Recommend "Tell a story" (animated, short movie)
     - Start with ESE 2020 vision animation but address the role of the observatory data systems
  2. Scenario storyboard with multiple user perspectives
     - Mission user/scientist with "direct" access
     - Graduate student researcher with access via value-added tools working within COTS
     - Commercial customers with access via value-added providers
     - Ordinary citizen making quiries (eg, directly from their tractor)
  3. Convey themes
     - SEEDS enables development of new products and services but does not make new products
     - Metric of success - commercial use of SEEDS standards, technology solutions
     - Demonstrate ability to connect data to sspecific user needs as needs evolve
  4. Different formats for different audiences
     - All formats under "considerations" are viable

Technology Infusion Barriers - Application Focus

- **What are the barriers to infusing new technology into current ESE data systems?**
  1. Different priorities between developers and operations
  2. Lack of modularity
  3. Systems not open by design
  4. COTS interdependencies
  5. Intellectual property issues- will the source be available? Bureaucratic release processes

Technology Infusion Strategies

- **What strategies can be used to ensure technologies are readily incorporated into operational ESE data systems?**
- **Considerations**
  1. Collaborative funding for technology customers to incorporate successful research prototypes into operational systems
  2. Sponsor activities connecting technology developers with technology customers
  3. Explore partnering opportunities
  4. Standardize key interfaces
- **(Note: topic not covered in the workshop due to time constraints)**

Next Steps: Capability Needs Information Roadmap